

Multi-Agent Approaches To Automated Annotation and Pathway Representation

Schmidt, Carl¹, Decker, Keith²

¹Department of Animal and Food Science and ²Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA

The move from a focus on standalone computer systems to the current situation of highly distributed, open, and dynamic systems has radically changed the nature of software development in the life sciences. Software agents are autonomous, problem-solving computational entities capable of effective operation in dynamic and open environments. Agents encompasses four properties:

1. Autonomy-agents are independent and function without constant programmer intervention.
2. Reactiveness-agents monitor their environment and respond to changes.
3. Pro-activeness-agents have goals that direct their behavior over long periods of time.
4. Social ability-agents must interact and cooperate with other software and human agents.

We are applying agent-based approaches to a number of bioinformatics problems, including automated sequence assembly and annotation, and pathway representation.

Our current tools for assembly and annotation, called BioMAS have been used to develop knowledgebases (chickKB) for chicken expressed sequence tags (ESTs) and for herpesvirus genomes (HVKB). BioMAS uses Multi-Agent Systems techniques to create a distributed open software system where largely independent, automated annotation and analysis programs operate on a database after an initial sequence assembly pipeline. These annotations include identifying genes by BLAST, protein motif and domain finding, transmembrane domain prediction, and electronic GO annotation using GODEL. By using the GO ontology as well as ontologies for protein domains and motifs, the automatically created database can be treated as a knowledgebase. This permits inferencing queries via a web interface, i.e. "Show me all glycoproteins in HSV1 with a phosphorylation motif and a transmembrane domain value ≥ 2 ". The Multi-Agent System approach gives us software tools and techniques for dealing with many of the problems in building such knowledgebases including dynamic, distributed primary source databases (Genbank), multiple analysis programs with inconsistent input/output formats, and dealing with large amounts of locally produced data that needs to be provided to the larger community via a web-accessible, ontology-driven query system.

Our current tools for signal transduction pathway representation and hypothesis generation rely on agent technologies for the representation of agent actions using artificial intelligence planning techniques. AI planning technology provides a modeling formalism for the task of biological pathway discovery, such that hypothetical pathways can be generated, queried and qualitatively simulated. The task of signal transduction pathway discovery is re-cast as a planning problem, one in which the initial and final states are known and cellular processes captured as abstract operators that modify the cellular environment. Thus, a valid plan that transforms the initial state into a goal state is a hypothetical pathway that prescribes the order of signaling events that must occur to effect the goal state. The planner is driven by data that is stored within a knowledge base and retrieved from heterogeneous sources (i.e. protein-protein interaction and literature mining) by a multi-agent information gathering system. We demonstrate the combined technology by translating the well-known EGF pathway into the planning formalism and deploying the Fast-Forward planner to reconstruct the pathway directly from the knowledge base.